

Elsevier L^AT_EX template^{*}

Elsevier¹

Radarweg 29, Amsterdam

Elsevier Inc^{a,b}, Global Customer Service^{b,}*

^a1600 John F Kennedy Boulevard, Philadelphia

^b360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted L^AT_EX manuscript.

Keywords: `elsarticle.cls`, L^AT_EX, Elsevier, template

2010 MSC: 00-01, 99-00

1. Introduction

Since the dawn of internet and world wide web, humanity has witnessed a degree of connection beyond reckoning. The proliferation of digital devices pervaded with various applications that account for almost all aspect of humanity, have created cyber communities that constantly mutate [1]; [2]. In a world where we have network infrastructures that can support up to 250Mbps of data transmission, and smart phones and IOT devices that can have processing power of up to 3 Ghz, data becomes ubiquitous, the quantum that lays the foundation of the nexus [3].

According to InternetLiveStates.com [4], only in one second, there are 9,878 tweets sent, 1,138 instagram photos uploaded, 3,117,720 emails sent, 99,738 Google searches made, and 94,144 Youtube videos viewed. That is, if it has

^{*}Fully documented templates are available in the elsarticle package on CTAN.

^{*}Corresponding author

Email address: `support@elsevier.com` (Global Customer Service)

URL: `www.elsevier.com` (Elsevier Inc)

¹Since 1880.

taken 5 second the read the preceding paragraph, during that time, 15,588,600 emails are sent.

15 Driven by the ambition to harness the power of this deluge of data, the term 'Big Data' (BD) was coined [5]. BD initially emerged to address the challenges associated with various characteristics of data such as velocity, variety, volume and variability [2]. BD is the practice of extracting patterns, theories, and predictions from a large set of structured, semi-structured, and unstructured
20 data for the purposes of business competitive advantage [6]; [7]. BD is a game-changing innovation, heralding the dawn of a new data-oriented industry.

Nonetheless, BD is not a magical wand that can enchant any business process. While a lot of opportunities exist in BD, subsuming an emergent and rather high-impacting technology like BD to current state of affairs in organi-
25 zations, is a daunting task. According to recent survey from Databricks, only 13% of the organizations excel at delivering on their data strategy [8]. Another survey by NewVantage Partners indicated that only 24% organization have successfully gone data-driven [9]. This survey also states that only 30% of organizations have a well established strategy for their big data endeavour. In
30 addition, surveys from McKinsey & Company ([10]) and Gartner ([11]) further support these numbers, which illuminates on the scarcity of successful big data implementations in the industry.

Among the challenges of data adoption perhaps the most highlighted are 'data engineering complexities', 'big data architecture', 'rapid technology change',
35 'lack of sufficient skilled data engineers', and 'organization's cultural challenges of becoming data-driven' [2];[12]. This focus of this study is on data engineering complexities and in specific big data architecture.

In the past, organization relied on a few technology giants to provide infrastructure and tools necessary for big data, while today there's a plethora of
40 choice from hundreds of providers covering different aspect of data ecosystem from ingestion, to logging, to stream processing, and to visualization [9]. Companies are tending more and more towards Cloud-native architectures for cost reduction, improved efficiency and new roles have been introduced such as chief

analytics officer (CAOs) and chief data officers (CDOs) to channel the organizational big data capabilities toward business value and competitive advantage [13].

So how can one embark on this rather sophisticated journey? what can be a good logical approach to absorb the ever-increasing complexity of big data systems? how can organizations build different stacks to handle data for various workloads such as machine learning (ML), business analytics, data engineering, and streaming?

We suggest that majority of the challenge discussed starts with data architecture [1]; [3]. The data ingestion, processing and consumption of different data workloads vary, and sometimes they don't go well together. A company that enacted a data lake and a data warehouse and tries to account for both ecosystems, can be dealing with immense complexity, which in turns impact data teams, which in turn can hinder innovation, create barriers and result in monumental lost.

Development and deployment of an efficacious big data system is only the beginning of a big data journey. As data sources increase, variety of data increases, number of data consumers increase, the data store gets confuscated, and this can introduce threats for scalability and maintainability of the system. This also implies that only a handful of hyper-specialized data engineers would understand the system internals, creating silos, and potential miscommunication.

Majority of these systems are developed on-premise as ad-hoc complicated solutions that do not adhere to the practices of software engineering and software architecture [14]; [15]. As the ecosystem grows and new technologies and data processing techniques are introduced, the software architect will have a harder time to come up with a solution that address the problem requirements.

This can potentially create grounds for an immature architecture that results in solutions that are hard to scale, hard to maintain, and raise high-entry blockades [3]. Since the approach of ad-hoc design to big data system development is not desirable and may leave many architects and data engineers in the dark,

75 novel data architectures that are designed specifically for BD are required. To
contribute to this goal, we explore the notion of reference architectures (RAs)
and present a distributed domain-driven software RA for big data systems.

2. Why reference architecture?

To justify why we have chosen reference architectures as the suitable artefact,
80 first we have to clarify two assumptions;

1. having a sound software architecture is essential to the successful devel-
opment and maintenance of software systems
2. there exist a sufficient body of knowledge in the field of software architec-
ture to support the development of an effective RA

85 One of the focal tenets of software architecture is that every system is devel-
oped to satisfy a business objective, and that the architecture of the system is a
bridge between abstract business goals to concrete final solutions [16]. While the
journey of big data can be quite challenging, the good news is that a software
RA can be designed, analyzed and documented incorporating best practices,
90 known techniques, and patterns that will support the achievement of the busi-
ness goals. In this way, the complexity can be absorbed, and made tractable.

Practitioners of complex systems, software engineers, and system designers
have been frequently using reference architectures to have a collective under-
standing of system components, functionalities, data-flows and patterns which
95 shape the overall qualities of system and help further adjust it to the business
objectives [17]; [18]. There is a fair amount of literature on reference architec-
tures, and whereas different authors definition may vary, they all share the same
tenets.

A reference architecture is amalgamation of architectural patterns, stan-
100 dards, software engineering techniques that bridge the problem domain to a
class of solutions. This artefact can be partially or completely instantiated and
prototyped in a particular business context together with other supporting arte-

fact to enable its use. RAs are often created from previous RAs and architecture [1].

105 *Usage.* Once the package is properly installed, you can use the document class *elsarticle* to create a manuscript. Please make sure that your manuscript follows the guidelines in the Guide for Authors of the relevant journal. It is not necessary to typeset your manuscript in exactly the same way as an article, unless you are submitting to a camera-ready copy (CRC) journal.

110 *Functionality.* The Elsevier article class is based on the standard article class and supports almost all of the functionality of that class. In addition, it features commands and options to format the

- document style
- baselineskip
- 115 • front matter
- keywords and MSC codes
- theorems, definitions and proofs
- labels of enumerations
- citation style and labeling.

120 **3. Front matter**

The author names and affiliations could be formatted in two ways:

- (1) Group the authors per affiliation.
- (2) Use footnotes to indicate the affiliations.

See the front matter of this document for examples. You are recommended to
125 conform your choice to the journal you are submitting to.

4. Bibliography styles

There are various bibliography styles available. You can select the style of your choice in the preamble of this document. These styles are Elsevier styles based on standard styles like Harvard and Vancouver. Please use BibTeX to
130 generate your bibliography and include DOIs whenever available.

Here are two sample references: [? ?].

References

References

- [1] P. Ataei, A. T. Litchfield, Big data reference architectures, a systematic
135 literature review.
- [2] B. B. Rad, P. Ataei, The big data ecosystem and its environs, International Journal of Computer Science and Network Security (IJCSNS) 17 (3) (2017) 38.
- [3] P. Ataei, A. Litchfield, Neomycelia: A software reference architecture for
140 big data systems, in: 2021 28th Asia-Pacific Software Engineering Conference (APSEC), IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 452–462. doi:10.1109/APSEC53868.2021.00052.
URL <https://doi.ieeecomputersociety.org/10.1109/APSEC53868.2021.00052>
- 145 [4] I. L. Stats, Internet live stats (2019).
- [5] M. Lycett, ‘datafication’: Making sense of (big) data in a complex world (2013).
- [6] B. B. Rada, P. Ataeib, Y. Khakbizc, N. Akbarzadehd, The hype of emerging technologies: Big data as a service.

- 150 [7] M. Huberty, Awaiting the second big data revolution: from digital noise to value creation, *Journal of Industry, Competition and Trade* 15 (1) (2015) 35–47.
- [8] M. technology review insights in partnership with Databricks, Building a high-performance data organization (2021).
 155 URL <https://databricks.com/p/whitepaper/mit-technology-review-insights-report>
- [9] N. Partners, Big data and ai executive survey 2021 (2021).
 URL https://www.supplychain247.com/paper/big_data_and_ai_executive_survey_2021/pragmadik
- 160 [10] M. Analytics, The age of analytics: competing in a data-driven world, Tech. rep., Technical report, San Francisco: McKinsey & Company (2016).
- [11] H. Nash, Cio survey 2015, Association with KPMG.
- [12] N. Singh, K.-H. Lai, M. Vejvar, T. Cheng, Big data technology: Challenges, prospects and realities, *IEEE Engineering Management Review*.
- 165 [13] B. B. Rad, P. Ataei, Evaluating major issues regarding reliability management for cloud-based applications, *IJCSNS* 17 (7) (2017) 168.
- [14] I. Gorton, J. Klein, Distribution, data, deployment, *STC 2015* (2015) 78.
- [15] S. Nadal, V. Herrero, O. Romero, A. Abelló, X. Franch, S. Vansummeren, D. Valerio, A software reference architecture for semantic-aware big data
 170 systems, *Information and software technology* 90 (2017) 75–92.
- [16] R. K. Len Bass, Dr. Paul Clements, *Software Architecture in Practice* (SEI Series in Software Engineering) 4th Edition, Addison-Wesley Professional; 4th edition, 2021.
- 175 [17] R. Cloutier, G. Muller, D. Verma, R. Nilchiani, E. Hole, M. Bone, The concept of reference architectures, *Systems Engineering* 13 (1) (2010) 14–27.

- [18] J. Kohler, T. Specht, Towards a secure, distributed, and reliable cloud-based reference architecture for big data in smart cities, in: *Big Data Analytics for Smart and Connected Cities*, IGI Global, 2019, pp. 38–70.